

Direct computation of diagnoses for ontology debugging^{*}

Kostyantyn Shchekotykhin, Philipp Fleiss, Patrick Rodler, and Gerhard Friedrich

Alpen-Adria Universität, Klagenfurt, 9020 Austria
firstname.lastname@aau.at

Abstract. Modern ontology debugging methods allow efficient identification and localization of faulty axioms defined by a user while developing an ontology. The ontology development process in this case is characterized by rather frequent and regular calls to a reasoner resulting in an early user awareness of modeling errors. In such a scenario an ontology usually includes only a small number of conflict sets, i.e. sets of axioms preserving the faults. This property allows efficient use of standard model-based diagnosis techniques based on the application of hitting set algorithms to a number of given conflict sets. However, in many use cases such as ontology alignment the ontologies might include many more conflict sets than in usual ontology development settings, thus making precomputation of conflict sets and consequently ontology diagnosis infeasible. In this paper we suggest a debugging approach based on a direct computation of diagnoses that omits calculation of conflict sets. Embedded in an ontology debugger, the proposed algorithm is able to identify diagnoses for an ontology which includes a large number of faults and for which application of standard diagnosis methods fails. The evaluation results show that the approach is practicable and is able to identify a fault in adequate time.

1 Introduction

Ontology development and maintenance relies on an ability of users to express their knowledge in form of logical axioms. However, the knowledge acquisition process might be problematic since a user can make a mistake in an axiom being modified or a correctly specified axiom can trigger a hidden bug in an ontology. These bugs might be of different nature and are caused by violation of *requirements* such as consistency of an ontology, satisfiability of classes, presence or absence of some entailments. In such scenarios as ontology matching the complexity of faults might be very high because multiple disagreements between ontological definitions and/or modeling problems can be triggered by axioms at once.

Ontology debugging tools [8,2,4] can simplify the development process by allowing their users specification of requirements to the intended (target) ontology. If some of the requirements are broken, i.e. an ontology \mathcal{O} is *faulty*, the debugging tool can compute a set of axioms $\mathcal{D} \subseteq \mathcal{O}$ called *diagnosis*. An expert should remove or modify at least all axioms of a diagnosis in order to be able to formulate the *target ontology* \mathcal{O}_t . Nevertheless, in real-world scenarios debugging tools can return a set of alternative diagnoses \mathbf{D} ,

^{*} The research project is funded by Austrian Science Fund (Project V-Know, contract 19996).

since it is quite hard for a user to specify such set of requirements that allows formulation of the target ontology \mathcal{O}_t only. Consequently, the user has to differentiate between multiple diagnoses $\mathcal{D} \in \mathbf{D}$ in order to find the *target diagnosis* \mathcal{D}_t , which application allows formulation of the intended ontology. Diagnosis discrimination methods [9,18] allow their users to reduce the number of diagnoses to be considered. The first approach presented in [9] uses a number of heuristics that rank diagnoses depending on the structure of axioms $ax_i \in \mathcal{D}$, usage in test cases, provenance information, etc. Only diagnoses with the highest ranks are returned to the user. A more sophisticated approach suggested in [18] identifies the target diagnosis by asking an oracle, like an expert or information extraction system, a sequence of questions: whether some axiom is entailed by the target ontology or not. Given an answer the algorithm removes all diagnoses that are inconsistent with it. Furthermore, the query is used to create an additional test case, which allows the search algorithm to prune the search space and reduce the number of diagnoses to be computed. Moreover, in order to speed up the computations the method approximates the set of all diagnoses with a set of n leading diagnoses, i.e. the n best diagnoses with respect to a given measure.

All the approaches listed above follow the standard model-based diagnosis approach [16] and compute diagnoses using minimal conflict sets CS , i.e. irreducible sets of axioms $ax_i \in \mathcal{O}$ that preserve violation of at least one requirement. The computation of the conflict sets can be done within a polynomial number of calls to the reasoner, e.g. by QUICKXPLAIN algorithm [7]. To identify a diagnosis of cardinality $|\mathcal{D}| = m$ the hitting set algorithm suggested in [16] requires computation of m minimal conflict sets. In the use cases when an ontology is generated by an ontology learning or matching system the number of minimal conflict sets m can be large, thus making the ontology debugging practically infeasible.

In this paper we present two algorithms INV-HS-TREE and INV-QUICKXPLAIN, which inverse the standard model-based approach to ontology debugging and compute diagnoses directly, rather than by means of minimal conflict sets. Thus, given some predefined number of leading diagnoses n the breadth-first search algorithm INV-HS-TREE executes the direct diagnosis algorithm INV-QUICKXPLAIN exactly n times. This property allows the new search approach to perform well when applied to ontologies with large number of conflicts. The evaluation shows that the direct computation of diagnoses allows to apply ontology debugging in the scenarios suggesting diagnosis of generated ontologies. The system based on INV-QUICKXPLAIN is able to compute diagnoses and identify the target one in the cases when common model-based diagnosis techniques fail. Moreover, the suggested algorithms are able to maintain a comparable or a slightly better performance in the cases that can be analyzed by both debugging strategies.

The remainder of the paper is organized as follows: Section 2 gives a brief introduction to the main notions of ontology debugging. The details of the suggested algorithms and their application are presented in Section 3. In Section 4 we provide evaluation results and conclude in Section 5.

2 Ontology debugging

Let us exemplify the ontology debugging process by the following use case:

Example 1. Consider an ontology \mathcal{O} with the terminology \mathcal{T} :

$$\begin{array}{lll} ax_1 : A \sqsubseteq B & ax_2 : B \sqsubseteq E & ax_3 : B \sqsubseteq D \sqcap \neg \exists s.C \\ ax_4 : C \sqsubseteq \neg(D \sqcup E) & ax_5 : D \sqsubseteq \neg B & \end{array}$$

and assertions $\mathcal{A} : \{A(w), A(v), s(v, w)\}$. Because of the axioms ax_3 and ax_5 the given terminology is incoherent and it includes two unsatisfiable classes A and B . Moreover, the assertions $A(w)$ and $A(v)$ make the ontology inconsistent.

Assume that the user is sure that the assertional axioms are correct, i.e. this part of the ontology is included to the *background knowledge* \mathcal{B} of the debugging system and, thus, cannot be considered as faulty in the debugging process. In this case, the debugger can identify the only irreducible set of axioms $CS' : \langle ax_3, ax_5 \rangle$ – minimal conflict set – preserving both inconsistency and incoherency of \mathcal{O} . Modification of at least one axiom of one of the minimal (irreducible) diagnoses $\mathcal{D}'_1 : [ax_3]$ or $\mathcal{D}'_2 : [ax_5]$ is required in order to restore both consistency and coherency of the ontology.

In some debugging systems, e.g. [18], the user can also provide *positive* P and *negative* N test cases, where each test case is a set of axioms that should be entailed (positive) and not entailed (negative) by the ontology resulting in application of the debugger. If in the example the user specifies

$$P = \{\{B(w)\}\} \text{ and } N = \{\{\neg C(w)\}\}$$

then the debugger returns another set of minimal conflicts sets:

$$CS_1 : \langle ax_1, ax_3 \rangle \quad CS_2 : \langle ax_2, ax_4 \rangle \quad CS_3 : \langle ax_3, ax_5 \rangle \quad CS_4 : \langle ax_3, ax_4 \rangle$$

and diagnoses:

$$\mathcal{D}_1 : [ax_3, ax_4] \quad \mathcal{D}_2 : [ax_2, ax_3] \quad \mathcal{D}_3 : [ax_1, ax_4, ax_5]$$

The reason is that both ontologies $\mathcal{O}'_1 = \mathcal{O} \setminus \mathcal{D}'_1$ and $\mathcal{O}'_2 = \mathcal{O} \setminus \mathcal{D}'_2$ resulting in application of the diagnoses \mathcal{D}'_1 and \mathcal{D}'_2 do not fulfill the test cases. For instance, \mathcal{O}'_2 is invalid since the axioms $\{ax_1, ax_3\} \subset \mathcal{O}'_2$ entail $\neg C(w)$, which must not be entailed.

Definition 1 (Target ontology). *The target ontology \mathcal{O}_t is a set of axioms that is characterized by a background knowledge \mathcal{B} , sets of positive P and negative N test cases. The target ontology \mathcal{O}_t should fulfill the following necessary requirements¹:*

- $\mathcal{O}_t \cup \mathcal{B}$ must be consistent and, optionally, coherent
- $\mathcal{O}_t \cup \mathcal{B} \models p \quad \forall p \in P$
- $\mathcal{O}_t \cup \mathcal{B} \not\models n \quad \forall n \in N$

The ontology \mathcal{O} is faulty with respect to a predefined \mathcal{B}, P and N iff \mathcal{O} does not fulfill the necessary requirements.

Definition 2 (Diagnosis problem instance). *$\langle \mathcal{O}, \mathcal{B}, P, N \rangle$ is a diagnosis problem instance, where \mathcal{O} is a faulty ontology, \mathcal{B} is a background theory, P is a set of test cases that must be entailed by the target ontology \mathcal{O}_t and N is a set of test cases that must not be entailed by \mathcal{O}_t . The instance is diagnosable if $\mathcal{B} \cup \bigcup_{p \in P} p$ is consistent (coherent) and $\mathcal{B} \cup \bigcup_{p \in P} p \not\models n$ for each $n \in N$.*

¹ In the following we assume that the user intends to formulate only one ontology. In the paper we refer to the intended ontology as the target ontology.

The ontology debugging approaches [4,8,18] can be applied to any knowledge representation language for which there is a sound and complete procedure for deciding whether an ontology entails an axiom or not. Moreover, the entailment relation \models must be extensive, monotone and idempotent.

Another important aspect of the ontology debugging systems comes from the model-based diagnosis techniques [16,10] that they are based on. The model-based diagnosis theory considers the modification operation as a sequence of add/delete operations and focuses only on deletion. That is, if an ontology includes faulty axiom then the simplest way to remove the fault is to remove the axiom. However, removing an axiom might be a too coarse modification, since the ontology can lose some of the entailments that must be preserved. Therefore, the model-based diagnosis takes also into account axioms (ontology extension) EX , which are added by the user to the ontology after removing all axioms of a diagnosis. Usually the set of axioms EX is either formulated by the user or generated by a learning system [11]. If EX is not empty then all axioms $ax_i \in EX$ are added to the set of positive test cases P , since each axiom ax_i must be entailed by the intended ontology \mathcal{O}_t .

Definition 3 (Diagnosis). For a diagnosis problem instance $\langle \mathcal{O}, \mathcal{B}, P, N \rangle$ a subset of the ontology axioms $\mathcal{D} \subset \mathcal{O}$ is a diagnosis iff $\mathcal{O}_t = (\mathcal{O} \setminus \mathcal{D})$ fulfills the requirements of Definition 1.

Due to computational complexity of the diagnosis problem, in practice the set of all diagnoses is approximated by the set of minimal diagnoses. For a diagnosis problem instance $\langle \mathcal{O}, \mathcal{B}, P, N \rangle$ a diagnosis \mathcal{D} is minimal iff there is no diagnosis \mathcal{D}' of the same instance such that $\mathcal{D}' \subset \mathcal{D}$.

The computation of minimal diagnoses in model-based approaches is done by means of conflict sets, which are used to constrain the search space.

Definition 4 (Conflict set). For a problem instance $\langle \mathcal{O}, \mathcal{B}, P, N \rangle$ a set of axiom $CS \subseteq \mathcal{O}$ is a conflict set iff one of the conditions holds:

- $CS \cup \mathcal{B} \cup \bigcup_{p \in P} p$ is inconsistent (incoherent) or
- $\exists n \in N$ such that $CS \cup \mathcal{B} \cup \bigcup_{p \in P} p \models n$

Just as for diagnoses, computation of conflict sets is reduced to computation of minimal conflict sets. A conflict set CS is *minimal* iff there is no conflict set CS' such that $CS' \subset CS$.

Computation of minimal conflict sets. In practice, the diagnosis systems use two types of strategies for computation of conflict sets, namely, brute-force [6,8] and divide-and-conquer [7]. The first strategy can be split into acquisition and minimization stages. During the acquisition stage the algorithm adds axioms of an ontology $\mathcal{O} \setminus \mathcal{B}$ to a buffer while a set of axioms in the buffer is not a conflict set. As soon as at least one conflict set is added to the buffer, the algorithm switches to the minimization stage. In this stage axioms are removed from the buffer such that the set of axioms in the buffer remains a conflict set after each deletion. The algorithm outputs a minimal conflict set or 'no conflicts'. In the worst case a brute force algorithm requires $O(m)$ calls to the reasoner, where m is the number of axioms in a faulty ontology. The algorithm implementing

divide-and-conquer strategy starts with a buffer containing all axioms of an ontology, i.e. the conflict set is in the buffer, and splits it into smaller and simpler sub-problems. The algorithm continues splitting until it identifies a sequence of sub-problems including only one axiom such that a set including all these axioms is a minimal conflict set. The divide-and-conquer algorithm requires in the worst case $O(k \log(\frac{m}{k}))$ calls to the reasoner, where k is the cardinality of a returned conflict set. Taking into account that in practice $k \ll m$, the divide-and-conquer strategy is preferred to the brute-force.

Identification of minimal diagnoses. The computation of minimal diagnoses in modern ontology debugging systems is implemented using the Reiter's Hitting Set HS-TREE algorithm [16,3]. The algorithm constructs a directed tree from root to the leaves, where each node nd is labeled either with a minimal conflict set $CS(nd)$ or \checkmark (consistent) or \times (pruned). The latter two labels indicate that the node is closed. Each edge outgoing from the open node nd is labeled with an element $s \in CS(nd)$. $HS(nd)$ is a set of edge labels on the path from the root to the node nd . Initially the algorithm creates an empty root node and adds it to the *queue*, thus, implementing a breadth-first search strategy. Until the queue is empty, the algorithm retrieves the first node nd from the queue and labels it with either:

1. \times if there is a node nd' , labeled with either \checkmark or \times , such that $H(nd') \subseteq H(nd)$ (pruning non-minimal paths), or
2. $CS(nd')$ if a node nd' exists such that its label $CS(nd') \cap H(nd) = \emptyset$ (reuse), or
3. CS if CS is a minimal conflict set computed for the diagnosis problem instance $\langle \mathcal{O} \setminus H(nd), \mathcal{B}, P, N \rangle$ by one of the algorithms mentioned above (compute), or
4. \checkmark (consistent).

The leaf nodes of a complete tree are either pruned (\times) or consistent (\checkmark) nodes. The set of labels $H(nd)$ of each consistent node nd corresponds to a minimal diagnosis. The minimality of the diagnoses is guaranteed due to the minimality of conflict sets, pruning rule and breadth-first search strategy. Moreover, because of the latter the minimal diagnoses are generated in order of increasing cardinality.

Diagnoses discrimination. In many real-world scenarios an ontology debugger can return a large number of diagnoses, thus, placing the burden of diagnosis discrimination on the user. Without an adequate tool support the user is often unable to understand the difference between the minimal diagnoses and to select an appropriate one. The diagnosis discrimination method suggested in [18] uses the fact that different ontologies, e.g. $\mathcal{O}_1 = \mathcal{O} \setminus \mathcal{D}_1$ and $\mathcal{O}_2 = \mathcal{O} \setminus \mathcal{D}_2$, resulting in the application of different diagnoses, entail different sets of axioms. Consequently, there exists a set of axioms Q such that $\mathcal{O}_1 \models Q$ and $\mathcal{O}_2 \not\models Q$. If such a set of axioms Q exists, it can be used as a query to some oracle such as the user or an information extraction system. If the oracle answers *yes* then the target ontology \mathcal{O}_t should entail Q and, hence, Q should be added to the set of positive test cases $P \cup \{Q\}$. Given the answer *no* the set of axioms is added to the negative test cases $N \cup \{Q\}$ to ensure that the target ontology does not entail Q . Thus, in the first case the set of axioms \mathcal{D}_2 can be removed from the set of diagnoses \mathbf{D} because \mathcal{D}_2 is not a diagnosis of the updated diagnosis problem instance $\langle \mathcal{O}, \mathcal{B}, P \cup \{Q\}, N \rangle$ according to Definition 3. Similarly, in the second case the set of axioms \mathcal{D}_1 is not a diagnosis of $\langle \mathcal{O}, \mathcal{B}, P, N \cup \{Q\} \rangle$.

However, many different queries might exist for the set of diagnoses $|\mathbf{D}| > 2$. In the extreme case there are $2^n - 2$ possible queries for a set of diagnoses including n elements. To select the best query the authors in [18] suggest two measures: SPLIT-IN-HALF and ENTROPY. The first measure is a greedy approach preferring the queries which allow to remove a half of the minimal diagnoses from \mathbf{D} , given an answer of an oracle. The second is an information-theoretic measure, which estimates the information gain for both outcomes of each query and returns the one that maximizes the information gain. The *prior fault probabilities* required for ENTROPY measure can be obtained from statistics of previous diagnosis sessions. For instance, if the user has problems with understanding of restrictions then the diagnosis logs will contain more repairs of axioms including restrictions. Consequently, the prior fault probabilities of axioms including restrictions should be higher. Given the fault probabilities of axioms, one can calculate prior fault probabilities of minimal diagnoses including these axioms as well as evaluate ENTROPY (see [18] for more details).

A general algorithm of the interactive ontology diagnosis process can be described as follows:

1. Generate a set of diagnoses \mathbf{D} including at most n diagnoses.
2. Compute a set of queries and select the best one according to some predefined measure.
3. Ask the oracle and, depending on the answer, add the query either to P or to N .
4. Update the set of diagnoses \mathbf{D} and remove the ones that do not comply with the newly acquired test case, according to the Definition 3.
5. Update the tree and repeat from Step 1 if the queue contains open nodes.
6. Return the set of diagnoses \mathbf{D} .

The resulting set of diagnoses \mathbf{D} includes only diagnoses that are not differentiable in terms of their entailments, but have some syntactical differences. The preferred diagnosis in this case should be selected by the user using some text differencing and comparison tool.

Note, that a similar idea can be found in [14] where authors use queries to an oracle to revise an ontology. Given a consistent and coherent ontology \mathcal{O} the system partitions it into two ontologies \mathcal{O}^{\models} and $\mathcal{O}^{\not\models}$ containing required and incorrect consequences correspondingly. The system can deal with inconsistent/incoherent ontologies if a union of all minimal conflict sets is put to the initial set of incorrect consequences $\mathcal{O}_0^{\not\models}$. Computation of the set $\mathcal{O}_0^{\not\models}$ requires application of an ontology debugger and is not addressed in [14].

Example 1, continued. Assume that the user mistakenly negated only the restriction in ax_3 instead of the whole description $B \sqsubseteq \neg(D \sqcap \exists s.C)$. Moreover, in ax_4 a disjunction was placed instead of conjunction because of a typo, i.e. $C \sqsubseteq \neg(D \sqcap E)$.

The interactive diagnosis process, illustrated in Fig. 1, applies the three techniques described above to find the target diagnosis $\mathcal{D}_t = [ax_3, ax_4]$. In the first iteration the system starts with the root node, which is labeled with $\langle ax_1, ax_3 \rangle$ – the first conflict returned by QUICKXPLAIN. Next, HS-TREE generates successor nodes and labels the edges leading to these nodes with corresponding axioms. The algorithm extends the search tree until two *leading minimal diagnoses* are computed. Given the set of minimal

of minimal conflict sets that can generate a large number of possible diagnoses. For instance, Transportation ontology (see [18]) includes only 9 minimal conflict sets that generate 1782 minimal diagnoses. In such case HS-TREE makes only 9 calls to QUICK-XPLAIN and then reuses the identified minimal conflicts to label all other nodes. The number of calls to the reasoner, which is the main “source of complexity”, is rather low and can be approximated by $9k \log \frac{n}{k} + |Nodes|$, where $|Nodes|$ is the cardinality of the set containing all nodes of the search tree, k is the maximum cardinality of all computed minimal conflict sets and n is the number of axiom in the faulty ontology. The combination of ontology debugging with diagnosis discrimination allows to reduce the number of calls to the reasoner, because often acquired test cases invalidate not only diagnoses that are already computed by HS-TREE, but also those that are not. All these factors together with such techniques as module extraction [17] make the application of ontology debuggers feasible in the described scenario.

However, in such applications as ontology matching or learning the number of minimal conflict sets can be much higher, because all axioms are generated at once. For instance, ontology alignments, identified by most of ontology matching systems in the last Ontology Alignment Evaluation Initiative (OAEI), are often incoherent and, in some cases, inconsistent [1]. The large number of minimal conflict set makes the application of ontology debugging problematic because of the large number of calls to the reasoner and the memory required by the breadth-first search algorithm. To overcome this problem we suggest a novel ontology debugging approach that computes diagnoses directly, i.e. without precomputation of minimal conflict sets.

3 Direct diagnosis of ontologies

The main idea behind the approach is to start with the set $\mathcal{D}_0 = \emptyset$ and extend it until such a subset of ontology axioms $\mathcal{D} \subseteq \mathcal{O}$ is found that \mathcal{D} is a minimal diagnosis with respect to the Definition 3. In the first step Algorithm 1 verifies the input data, i.e. if the input ontology \mathcal{O} , background knowledge \mathcal{B} , positive P and negative N test cases together constitute a valid diagnosis problem instance $\langle \mathcal{O}, \mathcal{B}, P, N \rangle$ (Definition 2). Thus it verifies: a) whether the background theory together with the positive and negative test cases is consistent; and b) if the ontology is faulty. In both cases INV-QUICKXPLAIN calls VERIFYREQUIREMENTS function that implements Definition 3 and tests if axioms in the set \mathcal{D} are a minimal diagnosis or not. The test function requires a reasoner that implements consistency/coherency checking (ISCONSISTENT) and allows to decide whether a set of axioms is entailed by the ontology (ENTAILS).

FINDDIAGNOSIS is the main function of algorithm which takes six arguments as input. The values of the arguments \mathcal{B} , \mathcal{O} and N remain constant during the recursion and are required only for verification of requirements. Whereas values of \mathcal{D} , Δ and \mathcal{O}_Δ are used to provide a set of axioms corresponding to the actual diagnosis and two diagnosis sub-problems on the next level of the recursion. The sub-problems are constructed during the execution FINDDIAGNOSIS by splitting a given diagnosis problem with SPLIT function. In the most of the implementations SPLIT simply partitions the set of axioms into two sets of equal cardinality. The algorithm continues to divide diagnosis problems (FINDDIAGNOSES line 12) until it identifies that the set \mathcal{D} is a diagnosis (line 7). In further iterations the algorithm minimizes the diagnosis by splitting it into sub-problems

Algorithm 1: INV-QUICKXPLAIN($\mathcal{O}, \mathcal{B}, P, N$)

Input: \mathcal{O} set of faulty axioms, \mathcal{B} set of background axioms, P set of positive test cases, N set of negative test cases
Output: a minimal diagnosis \mathcal{D}

```
1  $\mathcal{O}' \leftarrow \mathcal{O} \setminus \mathcal{B}$ ;  
2  $\mathcal{B}' \leftarrow \mathcal{B} \cup \bigcup_{p \in P} p$ ;  
3 if  $\neg \text{VERIFYREQUIREMENTS}(\mathcal{B}', \emptyset, \emptyset, N)$  then return 'inconsistent requirements';  
4 if  $\text{VERIFYREQUIREMENTS}(\mathcal{B}', \emptyset, \mathcal{O}', N)$  then return 'consistent';  
5 return  $\text{FINDDIAGNOSIS}(\mathcal{B}', \emptyset, \mathcal{O}', \mathcal{O}', N)$ ;  
  
6 function  $\text{FINDDIAGNOSIS}(\mathcal{B}, \mathcal{D}, \Delta, \mathcal{O}_\Delta, \mathcal{O}, N)$  returns a minimal diagnosis  $\mathcal{D}$   
7   if  $\Delta \neq \emptyset \wedge \text{VERIFYREQUIREMENTS}(\mathcal{B}, \mathcal{D}, \mathcal{O}, N)$  then return  $\emptyset$ ;  
8   if  $|\mathcal{O}| = 1$  then return  $\mathcal{O}$ ;  
9    $k \leftarrow \text{SPLIT}(\mathcal{O})$ ;  
10   $\mathcal{O}_1 \leftarrow \text{GETELEMENTS}(\mathcal{O}_\Delta, 1, k)$ ;  
11   $\mathcal{O}_2 \leftarrow \text{GETELEMENTS}(\mathcal{O}_\Delta, k - 1, |\mathcal{O}_\Delta|)$ ;  
12   $\mathcal{D}_2 \leftarrow \text{FINDDIAGNOSIS}(\mathcal{B}, \mathcal{D} \cup \mathcal{O}_1, \mathcal{O}_1, \mathcal{O}_2, \mathcal{O}, N)$ ;  
13   $\mathcal{D}_1 \leftarrow \text{FINDDIAGNOSIS}(\mathcal{B}, \mathcal{D} \cup \mathcal{O}_2, \mathcal{O}_2, \mathcal{O}_1, \mathcal{O}, N)$ ;  
14  return  $\mathcal{D}_1 \cup \mathcal{D}_2$ ;  
  
15 function  $\text{VERIFYREQUIREMENTS}(\mathcal{B}, \mathcal{D}, \mathcal{O}, N)$  returns true or false  
16    $\mathcal{O}' \leftarrow \mathcal{B} \cup (\mathcal{O} \setminus \mathcal{D})$ ;  
17    $c \leftarrow \text{ISCONSISTENT}(\mathcal{O}')$ ;  
18   if  $\neg c$  then return false;  
19   foreach  $n \in N$  do  $c \leftarrow c \wedge \text{ENTAILS}(\mathcal{O}', n)$ ;  
20   return  $c$ ;
```

of the form $\mathcal{D} = \mathcal{D}' \cup \mathcal{O}_\Delta$, where \mathcal{O}_Δ contains only one axiom. In the case when \mathcal{D} is a diagnosis and \mathcal{D}' is not, the algorithm decides that \mathcal{O}_Δ is a subset of the sought minimal diagnosis. Just as the original algorithm, INV-QUICKXPLAIN always terminates and returns a minimal diagnosis for a given diagnosis problem instance.

Example 1, continued. Let us look again at the ontology diagnosis example and show how a diagnosis is computed by INV-QUICKXPLAIN (see Fig. 2). The algorithm starts with an empty diagnosis $\mathcal{D} = \emptyset$ and \mathcal{O}_Δ containing all axioms of the problem. $\text{VERIFYREQUIREMENTS}$ returns *false* since the $\mathcal{B} \cup \mathcal{O} \setminus \emptyset$ is inconsistent. Therefore, the algorithm splits \mathcal{O}_Δ into $\{ax_1, ax_2\}$ and $\{ax_3, ax_4, ax_5\}$ and passes the sub-problem to the next level of recursion. Since, the set $\mathcal{D} = \{ax_1, ax_2\}$ is not a diagnosis, the ontology $\mathcal{B} \cup (\mathcal{O} \setminus \mathcal{D})$ is inconsistent and the problem in \mathcal{O}_Δ is split one more time. On the second level of recursion the set \mathcal{D} is a diagnosis, although not minimal. The function $\text{VERIFYREQUIREMENTS}$ returns *true* and the algorithm starts to analyze the found diagnosis. Therefore, it verifies whether the last extension of the set \mathcal{D} is a subset of a minimal diagnosis. Since, the extension includes only one axiom ax_3 and the extended set $\{ax_1, ax_2\}$ is not a diagnosis, the algorithm concludes that ax_3 is an element of the target diagnosis. The left-most branch of the recursion tree terminates and returns $\{ax_3\}$. This axiom is added to the set \mathcal{D} and the algorithm starts investigating whether the two axioms $\{ax_1, ax_2\}$ also belong to a minimal diagnosis. First, it tests the set

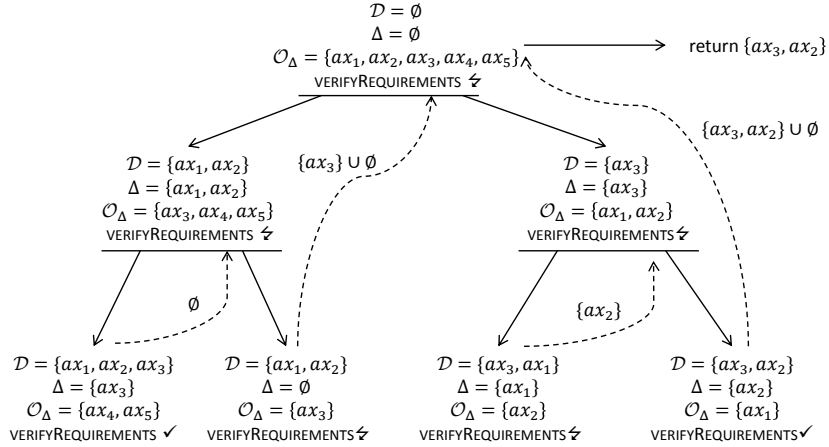


Fig. 2. Recursive calls of INV-HS-TREE for the diagnosis problem instance in Example 1. The background theory \mathcal{B} , original ontology \mathcal{O} and the set of negative test cases N remain constant and therefore are omitted. Solid arrows show recursive calls (line 12 – left and line 13 – right) and dashed arrows indicate returns.

$\{ax_3, ax_1\}$, which is not a diagnosis, and on the next iteration it identifies the correct result $\{ax_3, ax_2\}$.

INV-QUICKXPLAIN is a deterministic algorithm and returns the same minimal diagnosis if applied twice to a diagnosis problem instance. In order to obtain different diagnoses, the problem instance should be changed such that the INV-QUICKXPLAIN will identify the next diagnosis. Therefore, we suggest INV-HS-TREE, which is a modification of the HS-TREE algorithm presented in Section 2. The inverse algorithm labels each node nd of the tree with a minimal diagnosis $\mathcal{D}(nd)$. The rules 1, 2 and 4 of the original algorithm remain the same in INV-HS-TREE and the rule 3 is modified as follows:

3. The open node nd is labeled with \mathcal{D} if \mathcal{D} is a minimal diagnosis for the diagnosis problem instance $\langle \mathcal{O}, \mathcal{B} \cup H(nd), P, N \rangle$ computed by INV-QUICKXPLAIN (compute)

where $H(nd)$ is a set containing edge labels on the path from the root to nd . In this case elements of $H(nd)$ correspond to the axioms of minimal diagnoses that were used as labels of nodes on the path. Addition of an axiom ax_k of a minimal diagnosis \mathcal{D}_i to the background theory forces INV-QUICKXPLAIN to search for a minimal diagnosis that suggests the modification of any other axiom, except ax_k . That is, the diagnosis \mathcal{D}_i will not be rediscovered by the direct diagnosis algorithm.

A modified update procedure is another important feature of INV-HS-TREE. In the diagnosis discrimination settings the ontology debugger acquires new knowledge that can invalidate some of the diagnoses that are used as labels of the tree nodes. During the tree update INV-HS-TREE searches for the nodes with invalid labels. Given such a node, the algorithm removes its label and places it to the list of open nodes. Moreover, the algorithm removes all the nodes of a subtree originating from this node. After all

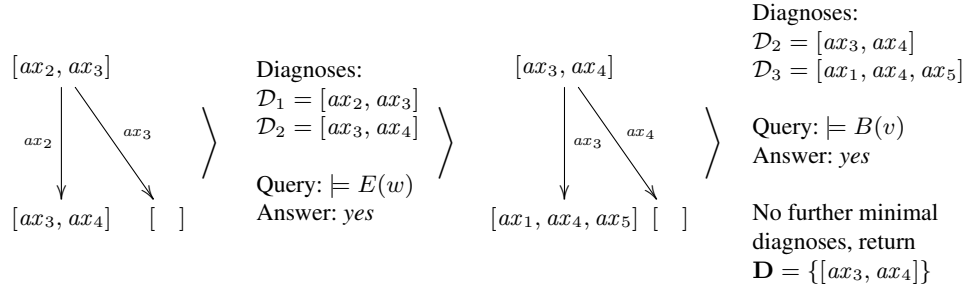


Fig. 3. Identification of the target diagnosis $[ax_3, ax_4]$ using direct diagnosis approach.

nodes with invalid labels are cleaned-up, the algorithm attempts to reconstruct the tree by reusing the remaining valid minimal diagnoses (rule 2, HS-TREE). Such aggressive pruning of the tree is feasible since a) the tree never contains more than n nodes that were computed with INV-QUICKXPLAIN and b) computation of a possible modification to the minimal diagnosis, that can restore its validity, requires invocation of INV-QUICKXPLAIN and, therefore, as hard as computation of a new diagnosis. Note also, that in a common diagnosis discrimination setting n is often set to a small number, e.g. 10, in order to achieve good responsiveness of the system. Consequently, in this settings the size of the tree will be small. The latter is another advantage of the direct method as it requires much less memory in comparison to a debugger based on the breadth-first strategy.

Example 1, continued. Applied to the sample diagnosis problem instance the direct ontology debugger computes two minimal diagnoses $[ax_2, ax_3]$ and $[ax_3, ax_4]$ in the first iteration (see Fig. 3). For these diagnoses the discrimination method identifies the query $E(w)$, which is answered *yes* by the oracle. The label of the root node in this case becomes invalid. Consequently, the algorithm removes the label of the root, deletes its subtree and places the root to the list of the open nodes. Next, according to the rule 2, the valid minimal diagnosis \mathcal{D}_2 is reused to label the root. Given a diagnosis problem instance $\langle \mathcal{O}, \mathcal{B} \cup \{ax_3\}, P, N \rangle$, INV-QUICKXPLAIN computes the last minimal diagnosis \mathcal{D}_3 . Given the positive answer to the query $B(v)$ the algorithm labels both open nodes with \checkmark and returns \mathcal{D}_2 as the result. Moreover, the labels on the edges of the tree correspond to the minimal conflicts $\langle ax_3 \rangle$ and $\langle ax_4 \rangle$ of the final diagnosis problem instance.

4 Evaluation

We evaluated the direct ontology debugging technique using aligned ontologies generated in the framework of OAEI 2011 [1]. These ontologies represent a real-world scenario in which a user generated ontology alignments by means of some (semi-)automatic tools. In such case the size of the minimal conflict sets and their configuration might be substantially different from the ones considered in the manual ontology development process, e.g. [8,18]. In the first experiment we demonstrate that INV-HS-TREE is able to identify minimal diagnoses in the cases when HS-TREE fails. The

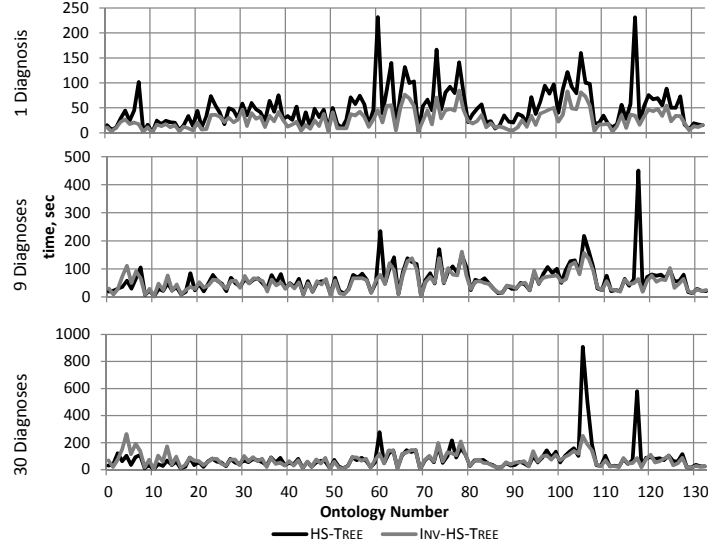


Fig. 4. Time required to compute 1, 9 and 30 diagnoses by HS-TREE and INV-HS-TREE for the Conference problem.

second test shows that the direct diagnosis approach is scalable and can be applied to ontologies including thousands of axioms.

The ontology matching problem can be formulated as follows: given two ontologies \mathcal{O}_i and \mathcal{O}_j , the goal of the ontology matching system is to identify a set of alignments M_{ij} . Each element of this set is a tuple $\langle x_i, x_j, r, v \rangle$, where $x_i \in Q(\mathcal{O}_i)$, $x_j \in Q(\mathcal{O}_j)$, r is a semantic relation and v is a confidence value. $Q(\mathcal{O})$ denotes a set of all matchable elements of an ontology \mathcal{O} such as classes or properties. The result of the ontology matching process is the aligned ontology $\mathcal{O}_{ij} = \mathcal{O}_i \cup M_{ij} \cup \mathcal{O}_j$. In the ontologies, used in both experiments, only classes and properties were considered as matchable elements and the set of relations was limited to $r \in \{\sqsubseteq, \equiv, \sqsupseteq\}$.

In the first experiment we applied the debugging technique to the set of aligned ontologies resulting from “Conference” set of problems, which is characterized by lower precision and recall of the applied systems (the best F-measure 0.65) in comparison, for instance, to the “Anatomy” problem (average F-measure 0.8)². The Conference test suite³ includes 286 ontology alignments generated by the 14 ontology matching systems. We tested all the ontologies of the suite and found that: a) 140 ontologies are consistent and coherent; b) 122 ontologies are incoherent; c) 26 ontologies are incon-

² see <http://oaei.ontologymatching.org/2011.5/results/index.html> for preliminary results of the evaluation based on reference alignments

³ All ontologies used in the evaluation can be downloaded from <http://code.google.com/p/rmbd/wiki/DirectDiagnosis>. The webpage contains also tables presenting detailed results of the experiment presented in Fig. 4.

Matcher	Ontology 1	Ontology 2	Time (ms) for		
			1 Diag	9 Diags	30 Diags
ldoa	cmt	edas	2540	6919	15470
csa	conference	edas	3868	14637	39741
ldoa	ekaw	iasted	10822	71820	229728
mappso	edas	iasted	11824	89707	293746
csa	edas	iasted	15439	134049	377361
csa	conference	ekaw	11257	31010	62823
ldoa	cmt	ekaw	5602	19730	42284
ldoa	conference	confof	8291	23576	48062
ldoa	conference	ekaw	7926	27324	56988
mappso	conference	ekaw	11394	33763	70469
mappso	confof	ekaw	9422	25921	55667
optima	conference	ekaw	11108	29837	62131
optima	confof	ekaw	7424	22506	44528

Table 1. Ontologies diagnosable only with INV-HS-TREE.

sistent; and in 8 cases HerMiT [13] was unable to finish the classification in two hours⁴. The results show that only two systems CODI and MaasMtch out of 14 were able to generate consistent and coherent alignments. This observation confirms the importance of high-performance ontology debugging methods.

The 146 ontologies of the cases b) and c) were analyzed with both HS-TREE and INV-HS-TREE. For each of the ontologies the system computed 1, 9 and 30 leading minimal diagnoses. The results of the experiment presented in Fig.4 show that for 133 ontologies out of 146 both approaches were able to compute the required amount of diagnoses. In the experiment where only 1 diagnosis was requested, the direct approach outperforms the HS-TREE as it was expected. In the next two experiments the time difference between the approaches decreases. However, the direct approach was able to avoid a rapid increase of computation time for very hard cases.

In the 13 cases presented in Table 1 the HS-TREE was unable to find all requested diagnoses in each experiment. Within 2 hours the algorithm calculated only 1 diagnosis for *csa-conference-ekaw* and for *ldoa-conference-confof* it was able to find 1 and 9 diagnoses. The results of the INV-HS-TREE are comparable with the presented in the Fig. 4. This experiment shows that the direct diagnosis is a stable and practically applicable method even in the cases when an ontology matching system outputs results of only moderate quality.

Moreover, in the first experiment we evaluated the efficiency of the interactive direct debugging approach applied to the cases listed in Table 1. In order to select the target diagnosis we searched for all possible minimal diagnoses of the following diagnosis problem instance $\langle M_f, \mathcal{O}_i \cup \mathcal{O}_j \cup M_t, \emptyset, \emptyset \rangle$, where M_f and M_t are the sets of *false* and *true* positive alignments. Both sets can be computed from the set of correct alignments M_c , provided by the organizers of OAEI 2011, and the set M_{ij} generated by a ontology matching system as $M_f = M_{ij} \setminus M_c$ and $M_t = M_{ij} \cap M_c$. From this

⁴ The tests were executed on a Core-i7 (3930K) 3.2Ghz, 32GB RAM and with Ubuntu Server 11.04, Java 6 and HetmiT 1.3.6 installed.

Matcher	Ontology 1	Ontology 2	Scoring	Time (ms)	#Query	React (ms)	#CC	CC (ms)
ldoa	conference	confof	ENT	11624	6	1473	430	3
ldoa	conference	confof	SPL	11271	7	1551	365	4
ldoa	cmt	ekaw	ENT	48581	21	2223	603	16
ldoa	cmt	ekaw	SPL	139077	49	2778	609	54
mappso	confof	ekaw	ENT	9987	5	1876	341	7
mappso	confof	ekaw	SPL	31567	13	2338	392	21
optima	conference	ekaw	ENT	16763	5	2553	553	8
optima	conference	ekaw	SPL	16055	8	1900	343	12
optima	confof	ekaw	ENT	23958	20	1137	313	14
optima	confof	ekaw	SPL	17551	10	1698	501	6
ldoa	conference	ekaw	ENT	56699	35	1458	253	53
ldoa	conference	ekaw	SPL	25532	9	2742	411	16
csa	conference	ekaw	ENT	6749	2	2794	499	3
csa	conference	ekaw	SPL	22718	8	2674	345	20
mappso	conference	ekaw	ENT	27451	13	1859	274	28
mappso	conference	ekaw	SPL	70986	16	4152	519	41
ldoa	cmt	edas	ENT	24742	22	1037	303	8
ldoa	cmt	edas	SPL	11206	7	1366	455	2
csa	conference	edas	ENT	18449	6	2736	419	5
csa	conference	edas	SPL	240804	37	6277	859	36
csa	edas	iasted	ENT	1744615	3	349247	1021	1333
csa	edas	iasted	SPL	7751914	8	795497	577	11497
ldoa	ekaw	iasted	ENT	23871492	10	1885975	287	72607
ldoa	ekaw	iasted	SPL	20448978	9	2100123	517	37156
mappso	edas	iasted	ENT	18400292	5	2028276	723	17844
mappso	edas	iasted	SPL	159298994	11	13116596	698	213210

Table 2. Diagnosis discrimination using direct ontology debugging. **Scoring** stands for query selection strategy, **react** system reaction time between queries, **#CC** number of consistency checks, **CC** gives average time needed for one consistency check.

set of diagnoses we choose one diagnosis at random as the target. In the experiment the prior fault probabilities of diagnoses were assumed to be $1 - v$, where v is the confidence value of the ontology matching system that the alignment is correct. Moreover, all axioms of both ontologies \mathcal{O}_i and \mathcal{O}_j were assumed to be correct and were assigned small probabilities.

The results presented in Table 2 were computed using both split-in-half (SPL) and entropy measure (ENT) for query selection for the diagnosis problem instance $\langle M_{ij}, \mathcal{O}_i \cup \mathcal{O}_j, \emptyset, \emptyset \rangle$. The entropy measure was able to solve the problem more efficiently because it is able to use information provided by the ontology matcher in terms on confidence values. The experiment shows also that efficiency any debugging methods depends highly on the ability of the underlying reasoner to classify an ontology. Note that the comparison of the suggested debugging technique with the ones build-in to such ontology matching systems as CODI [15] or LogMap [5] is inappropriate, since all these systems use greedy diagnosis techniques (e.g. [12]), whereas the method presented in this paper is complete. However, the results presented in Table 1 as well

Matcher	Scoring	INV-HS-TREE	HS-TREE
AgrMaker	ENT	19.62	20.833
AgrMaker	SPL	36.035	36.034
GOMMA-bk	ENT	18.343	14.472
GOMMA-bk	SPL	18.946	19.512
GOMMA-nobk	ENT	18.261	14.255
GOMMA-nobk	SPL	18.738	19.473
Lily	ENT	78.537	82.524
Lily	SPL	82.944	115.242
LogMap	ENT	6.595	13.406
LogMap	SPL	6.607	15.133
LogMapLt	ENT	14.847	12.888
LogMapLt	SPL	15.589	17.45
MapSSS	ENT	81.064	56.169
MapSSS	SPL	88.316	77.585

Table 3. Scalability test for INV-HS-TREE, time given in seconds.

as in Fig. 4 indicate that the suggested approach can find one minimal diagnosis in 25 seconds on average, which is comparable with the time of the mentioned systems.

In the second evaluation scenario we applied the direct method to unsatisfiable ontologies, generated for the Anatomy problem. The source ontologies \mathcal{O}_1 and \mathcal{O}_2 include 11545 and 4838 axioms correspondingly, whereas the size of the alignments varies between 1147 and 1461 axioms. The diagnosis selection process was performed in the same way as in the first experiment, i.e. we selected randomly one of the diagnoses of the instance $\langle M_f, \mathcal{O}_1 \cup \mathcal{O}_2 \cup M_t, \emptyset, \emptyset \rangle$. The tests were performed for the problem instance diagnosis $\langle M_{12}, \mathcal{O}_1 \cup \mathcal{O}_2, \emptyset, \emptyset \rangle$ for 7 of the 12 systems. We excluded the results of CODI, CSA, MaasMtch, MapEVO and Aroma, because CODI produced coherent alignments and the output of the other systems was not classifiable within 2 hours. The results of the experiment show that the target diagnosis can be computed within 40 second in an average case. Moreover, INV-HS-TREE slightly outperformed HS-TREE.

5 Conclusions

In this paper we present an approach to direct computation of diagnoses for ontology debugging. By avoiding computation of conflict sets, the algorithms suggested in the paper are able to diagnose the ontologies for which a common model-based diagnosis technique fails. Moreover, we show that the method can also be used with diagnosis discrimination algorithms, thus, allowing interactive ontology debugging. The experimental results presented in the paper indicate that the performance of a system using the direct computation of diagnoses is either comparable with or outperforms the existing approach in the settings, when faulty ontologies are generated by ontology matching or learning systems. The scalability of the algorithms was demonstrated on a set of big ontologies including thousands of axioms.

References

1. Euzenat, J., Ferrara, A., van Hage, W.R., Hollink, L., Meilicke, C., Nikolov, A., Ritze, D., Scharffe, F., Shvaiko, P., Stuckenschmidt, H., Sváb-Zamazal, O., dos Santos, C.T.: Final results of the Ontology Alignment Evaluation Initiative 2011. In: Proceedings of the 6th International Workshop on Ontology Matching. pp. 1–29. CEUR-WS.org (2011)
2. Friedrich, G., Shchekotykhin, K.: A General Diagnosis Method for Ontologies. In: Proceedings of 4th International Semantic Web Conference. pp. 232–246. Springer (2005)
3. Greiner, R., Smith, B., Wilkerson, R.: A correction to the algorithm in Reiter’s theory of diagnosis. *Artificial Intelligence* 41(1), 79–88 (1989)
4. Horridge, M., Parsia, B., Sattler, U.: Laconic and Precise Justifications in OWL. In: Proceedings of the 7th International Semantic Web Conference. pp. 323–338. Springer (2008)
5. Jiménez-Ruiz, E., Grau, B.C.: Logmap: Logic-based and scalable ontology matching. In: International Semantic Web Conference (ISWC). pp. 273–288. Springer (2011)
6. Junker, U.: QUICKXPLAIN: Conflict Detection for Arbitrary Constraint Propagation Algorithms. In: IJCAI’01 Workshop on Modelling and Solving Problems with Constraints (2001)
7. Junker, U.: QUICKXPLAIN: Preferred Explanations and Relaxations for Over-Constrained Problems. In: Proceedings of the Nineteenth National Conference on Artificial Intelligence. vol. 3, pp. 167–172 (2004)
8. Kalyanpur, A., Parsia, B., Horridge, M., Sirin, E.: Finding all Justifications of OWL DL Entailments. In: Proceedings of 6th International Semantic Web Conference. pp. 267–280. Berlin, Heidelberg (2007)
9. Kalyanpur, A., Parsia, B., Sirin, E., Cuenca-Grau, B.: Repairing Unsatisfiable Concepts in OWL Ontologies. In: The Semantic Web: Research and Applications, 3rd European Semantic Web Conference. pp. 170–184 (2006)
10. de Kleer, J., Williams, B.C.: Diagnosing multiple faults. *Artificial Intelligence* 32(1), 97–130 (Apr 1987)
11. Lehmann, J., Bühmann, L.: ORE - A Tool for Repairing and Enriching Knowledge Bases. In: The Semantic Web - ISWC 2010 - 9th International Semantic Web Conference. pp. 177–193 (2010)
12. Meilicke, C., Stuckenschmidt, H.: An Efficient Method for Computing Alignment Diagnoses. In: Proceedings of the 3rd International Conference on Web Reasoning and Rule Systems. pp. 182–196 (2009)
13. Motik, B., Shearer, R., Horrocks, I.: Hypertableau Reasoning for Description Logics. *Journal of Artificial Intelligence Research* 36(1), 165–228 (2009)
14. Nikitina, N., Rudolph, S., Glimm, B.: Interactive Ontology Revision. *Web Semantics: Science, Services and Agents on the World Wide Web* 12-13, 118–130 (2012)
15. Noessner, J., Niepert, M., Meilicke, C., Stuckenschmidt, H.: Leveraging Terminological Structure for Object Reconciliation. *The Semantic Web: Research and Applications* pp. 334–348 (2010)
16. Reiter, R.: A Theory of Diagnosis from First Principles. *Artificial Intelligence* 32(1), 57–95 (1987)
17. Sattler, U., Schneider, T., Zakharyashev, M.: Which Kind of Module Should I Extract? In: Proceedings of the 22nd International Workshop on Description Logics. CEUR Workshop Proceedings, vol. 477. CEUR-WS.org (2009)
18. Shchekotykhin, K., Friedrich, G., Fleiss, P., Rodler, P.: Interactive ontology debugging : two query strategies for efficient fault localization. *Web Semantics: Science, Services and Agents on the World Wide Web* 12-13, 88–103 (2012)